Chem. Senses 30: 547–557, 2005

doi:10.1093/chemse/bji048
Advance Access publication August 3, 2005

Building a Tree of Knowledge: Analysis of Bitter Molecules

Sarah Rodgers, Johanneke Busch, Hans Peters and Elly Christ-Hazelhof

Unilever Food and Health Research Institute, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

Correspondence to be sent to: S. Rodgers, Unilever Food and Health Research Institute, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands. e-mail: sarah.rodgers@unilever.com

Abstract

A phylogenetic-like tree of structural fragments has been constructed to extract useful insights from a structural database of bitter molecules. The tree of structural fragments summarizes the substructural groups present in the molecules from the bitter database. These structural fragments are compared with a large number of random molecules to highlight substructures specific to bitter molecules. This organization of the structures enabled the detection of structure–activity relationships for the bitter molecules through the construction of R-tables. Key structural groups, able to distinguish between bitter and random molecules, were identified through an analysis of the tree. This information can be used to further understand which structural components are involved in producing a bitter taste.

Key words: bitter, molecular similarity, phylogenetic-like tree, R-tables, structure–activity relationships, taste

Introduction

Many functional food ingredients, compounds that have beneficial effects on health, have been found to elicit a bitter taste. This is clearly an undesirable feature in most foodstuffs and is likely to lead to rejection by the consumer. Bitter taste perception is thought to have evolved as a mechanism to prevent the ingestion of potential toxins. However, many bitter compounds, such as plant-based phenols and flavonoids, can be beneficial in small amounts despite being toxic at higher concentrations (Ames *et al*., 1990). To reap the health benefits of these compounds the bitter taste must be managed in a way that is acceptable to the consumer.

A large range of molecules is known to confer a bitter taste, including organic molecules, peptides, inorganic ions and salts. Of the five basic tastes: sweet, bitter, sour, salty and umami, bitter appears to be the most complex (Behrens *et al.*, 2004). Sweet-tasting molecules require a hydrogen bond donor group and hydrogen bond acceptor group (Shallenberger and Acree, 1967), often in addition to a hydrophobic group (Kier, 1972), in order to exert a sweet taste. Bitter molecules can be structurally very similar to sweet molecules with only minor alterations rendering previously sweet molecules bitter (Drewnowski, 2001); for example, aspartame is sweet and one of its stereoisomers is bitter. Bitterness and sweetness are clearly closely related, but, partly as a consequence of such a diverse range of molecules being involved, bitterness is poorly understood.

Sour and salty taste is exerted through specialized membrane ion channels (Chadwick *et al.*, 1993), whereas sweet, bitter and umami taste operate via G protein-coupled receptors. It is currently agreed that 25 human G protein-coupled receptors mediate bitter taste (TAS2Rs); these have been identified through molecular cloning and functional studies (Adler *et al*., 2000; Chandrashekar *et al*., 2000). A large number of receptors is required to accommodate the diverse range of bitter molecules; however, there is very little information detailing which molecules are associated with which receptors. This is due in part to the difficulties involved in functionally expressing the bitter receptors in cultured mammalian cell lines (Pronin *et al*., 2004). In addition, there is no distinction in the signal produced from the different bitter receptors, preventing people from being able to discriminate between different bitter molecules.

The number and diversity of bitter tastants indicate that several mechanisms of bitter transduction may be involved. It has been suggested that in addition to functioning as monomers, bitter receptors may operate as heteromeric receptors in order to accommodate this diversity (Andres-Barquin and Conte, 2004). It is also possible that not all bitter tastants function via the G protein-coupled receptors; lipophilic bitter compounds and bitter salts may activate intracellular signals via alternative mechanisms (Andres-Barquin and Conte, 2004).

© The Author 2005. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oupjournals.org

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

As so little is currently understood about the relationship between bitterness and molecular structure, the bitter molecules are considered collectively here to identify distinguishing features on the bitter tastants. This may reveal structure–activity relationships and functional groups that can be investigated further. Molecular informatics techniques that identify structurally similar molecules are based on the *similar property principle* (Johnson and Maggiora, 1990). This states that structurally similar molecules tend to have similar properties, both physical and biological. Thus, grouping molecules according to their chemical structure allows identification of molecules that may share the same activity, in this case binding to specific bitter receptors.

The structures were studied through the construction of a phylogenetic-like tree (PGLT) (Nicolaou *et al.*, 2002), a tree of structural fragments that summarizes the structural features present in a database of molecules. This identifies the structural features on bitter molecules that differentiate them from non-bitter molecules; these are likely to be those features that are involved in binding to the bitter receptor. Although not previously used on taste molecules, PGLTs have been effectively employed to obtain structure–activity information from anti-HIV (Tamura *et al.*, 2002) and bacterial mutagenicity (Bacha *et al.*, 2002) data sets. In the present study the ability of the PGLT method to represent the database of bitter molecules to facilitate the identification of structure–activity relationships will be explored.

Previous research has included a general mapping of the three-dimensional arrangement of the bitter receptor active site (Tancredi *et al.*, 1979). Since the identification of the group of receptors responsible for mediating bitterness, many binding assays have been developed to identify ligands specific to the different bitter receptors (Bufe *et al.*, 2002; Behrens *et al.*, 2004). Other groups have clustered molecules according to subject's sensitivities to them (Delwiche *et al.*, 2001; Keast and Breslin, 2002); this is based on the assumption that if a number of subjects are similarly sensitive to a pair of molecules and another group of subjects is similarly insensitive, then the two molecules may share a common binding mechanism. Froloff *et al.* (1996) used a molecular modelling approach to identify common binding motifs for 14 different taste compounds (bitter and sweet). The structural distance between fragments representing the compounds was correlated with the difference in taste between the compounds.

An alternative molecular modelling approach is described here to identify structure–activity relationships for bitterness. A much larger data set of molecules will be analysed and a different method than that of Froloff *et al.* (1996) is used; by considering the structures of bitter molecules collectively, and comparing these structures with random molecules, key structural features associated with bitterness have been identified. This may help in our understanding of how these structures produce the bitter taste, how we can manage this (for example with masking), and how bitterness in compounds can be predicted.

Materials and methods

A database of bitter molecules was constructed from an extensive scan of scientific literature and patents, by conducting searches on bitterness in the Food Science and Technology Abstracts (FSTA, 2003), BIOSIS (BIOSIS, 2003), Derwent World Patents Index (WPIDS, 2003) and databases of internal reports. This led to over 2300 articles and patents being read to identify bitter molecules; the chemical name, structure and sensory quality of the identified compounds were put into the database. A molecule was considered bitter when it was reported as bitter. In the subsequent structural analyses, threshold data relating to the molecules was not considered due to the range of methods from the different sources of this data. Where possible, structures of bitter compounds were extracted from SciFinder (SciFinder, 2004). Alternatively, structures were searched for within the *Dictionary of Natural Products* (DNP, 2004) and the *World Drug Index* (WDI, 2004), or they were constructed using ChemDraw (ChemDraw, 2001). All structures were stored as SD files. Duplicates, salts and ions were removed from the data set, resulting in 833 bitter molecules to be used as the training set. A summary of the main structural groups in the training set is given in Figure 1.

The database also contains a small number of lipid molecules, aglycones, ureas, and hop alpha and beta acids. As a number of peptides were extracted from the literature they were also included in this study. However, only those peptides containing less than five amino acids were used, as peptides with more amino acids are considered too large for comparisons with small molecules.

Two-dimensional molecular fingerprints were created for these 833 bitter molecules to enable cluster analysis. A fingerprint is a binary string describing the presence or absence of substructural features and atoms in a given molecule. Molecular fingerprints are commonly used to provide an effective way of storing and representing molecular structures computationally; Unity fingerprints (Unity, 2004) were utilized here.

Cluster analysis was employed to organize the database of bitter molecules into distinct groups based on their chemical structure, such that molecules in the same cluster are structurally similar and molecules in different clusters are structurally dissimilar. The clustering algorithm *k*-means (Hartigan and Wong, 1979) was used to partition the database of bitter molecules into 30 clusters. The resulting clusters were examined both manually through visual inspection of the structures and by calculating the degree of structural similarity between cluster members. The similarity measure employed was the Tanimoto coefficient (Jaccard, 1901), a coefficient often used for the analysis of chemical structure data.

The phylogenetic-like tree (PGLT) is produced by generating a number of maximal common substructures (MCSs) from sub-groups, or clusters, of molecules within the

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

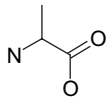
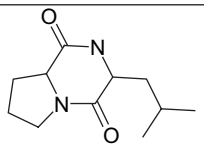
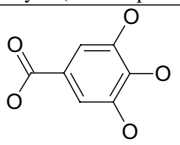
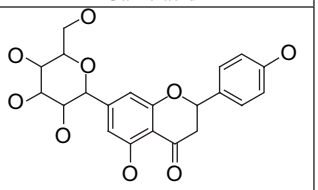
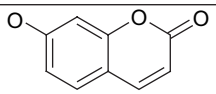
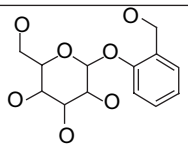
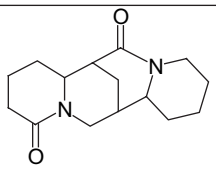
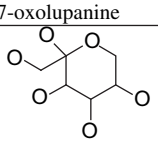
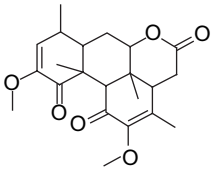
Structural Class	Example Structure
Amino acid (Meiselman and Halpern, 1970; Suzuki *et al.,* 2002)	 Alanine
Diketopiperazine (Ishibashi, *et al*., 1988)	 Cyclo(leucine-proline)
Flavonoid (Robichaud and Noble, 1990; Lee *et al.,* 1999)	 Gallic acid
Glucoside (Rouseff, 1980; Drewnowski and Gomez-Carneros, 2000)	 Prunin
Lactone (Berry and Tatum, 1986; Price *et al*., 1990)	 7-hydroxycoumarin
Pyranoside (Bufe *et al.,* 2002)	 Salicin
Quinolizidine alkaloid (Hatzold *et al*., 1983; Ruiz and Sotelo, 2001)	 17-oxolupanine
Sugar, sugar derivative (Ruiz-Avalia *et al*., 2000; Schiffman *et al*., 1995)	 Fructose
Terpene (Belitz and Wieser, 1985)	 Quassin

Figure 1 Summary of main structural groups included in dataset.

database. A MCS is the largest substructure common to a group of structurally related molecules. Each of the substructures is organized into a tree, where at each node there is a substructure and hierarchical relationships exist between the nodes (and therefore substructures) of the tree. All molecules in the database are tested to determine whether they contain a particular substructure; those that do are added to the relevant node. The end result is a tree of molecular substructures that summarizes the structural components from the database. Thus, the target is for all of the chemical groups present in the database of bitter molecules to be represented by the PGLT. Examination of the different structural groupings can provide information about functional groups associated with bitterness.

A simple example of a PGLT is given in Figure 2. The nodes are labelled 1, 1.1 and 1.2, the accompanying substructures are given in the boxes, and three of the molecules from the database are labelled A, B and C. Nodes 1.1 and 1.2 are children of node 1, as they have been created from this parent node; all members of nodes 1.1 and 1.2 are also members of node 1. Molecules A, B and C are tested to determine whether they contain the substructure of node 1: molecules A and B do and are assigned to node 1. Molecule C does not contain this substructure and thus is not added to node 1. Molecules A and B are then tested to determine whether they contain the substructures of the child nodes of node 1. Only the parent node members are tested on child nodes so molecule C is not included here. Molecule A contains the substructure of node 1.1 and node 1.2, and thus is assigned to both of these nodes; molecule B has the substructure of node 1.2, and thus is assigned membership in node 1.2 only. There is no limit on membership in the PGLT;

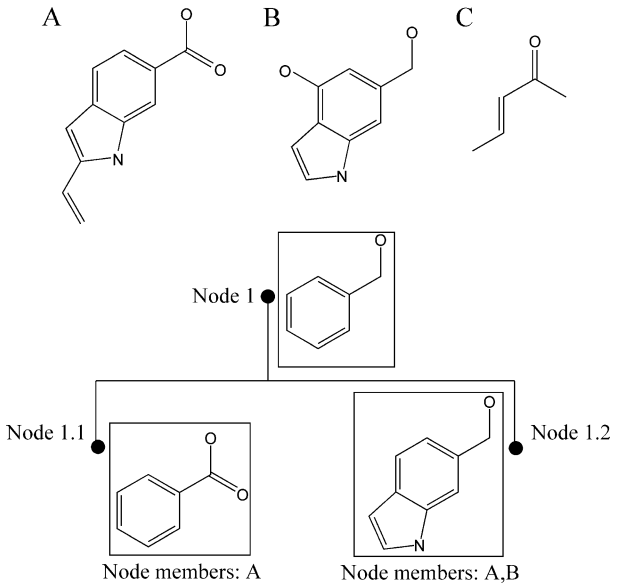


Figure 2 Example of a PGLT with node members.

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

molecules are assigned to all of the nodes with which they share a substructure.

The method followed to create the PGLT for bitter molecules is based on that described by Nicolau *et al.* (2002) and can be summarized as follows:

1. Perform cluster analysis on the whole data set (node members in subsequent runs).
2. Analyse clusters, exclude any with low internal similarity (using the Tanimoto coefficient).
3. Calculate the maximal common substructure (MCS) for each cluster; examine these and exclude any that are too similar to any already produced. These are the nodes' substructures.
4. Filter through molecules from the whole data set (parent node in subsequent runs) to find those that contain the MCS of each new node, assign membership to those that do.
5. All nodes with more than 10 members are considered 'GO' nodes—further cluster analysis is performed (Steps 1–4) to produce more child nodes. Nodes with fewer than 10 members are terminal nodes, or 'leaves'.
6. Once all of the nodes have been created, post-filter the data set of random molecules.

MCS identification and subsequent substructural filtering with the database of molecules was performed using PipeLine Pilot (PipeLine Pilot, 2004). The inclusion of random molecules in the PGLT enables the identification of those substructures that are specific to bitter molecules. A database of non-bitter molecules would have been more suitable here, but as bitterness is poorly defined, there are no large data sets of non-bitter molecules. For this study a large database is required for the post-filtering stage; the *Dictionary of Natural Products* (DNP, 2004) was used to fulfil this role. Molecules were selected randomly from the DNP and all duplicates, salts, ions and known bitter molecules removed.

Results

Filtering a set of mainly random molecules through the nodes in the tree reveals information about which structural fragments are common to both bitter and random molecules and, more importantly, which substructures are more specific to bitter molecules. Of the 31 971 DNP molecules, 23 179 contained at least one of the substructures from the tree. The tree also contains a total of 758 bitter molecules; thus 3.2% of the molecules with membership in the tree nodes are from the bitter database. By considering the proportion of known bitter molecules in each of the nodes of the tree and comparing this to the proportion in the tree as a whole, important and significant substructures can be identified. Nodes considered in this study to be significant are those containing >20% bitter molecules, a six-factor significance level (when compared to the proportion of bitter molecules present in the whole tree). A proportion of bitter molecules in a given node greater than this threshold indicates that the corresponding substructure has effectively differentiated between bitter and random compounds. When examining the proportions in each of the nodes, the total membership must also be considered as the nodes differ largely in size. Nodes containing 2 or 100 molecules may have the same proportion of bitter molecules, but the significance of the node will be very different.

Each node of the PGLT contains a group-defining common substructure and list of bitter and random molecules that contain this substructure. The tree contains a total of 116 nodes, that is 116 different substructures that are present in the dataset of bitter molecules. Only those nodes containing more than eight members are considered here; this leaves a total of 93 nodes. The substructures representing those nodes containing a significant proportion of bitter molecules are given in Figure 3. Spatial constraints prevent the inclusion of all of the nodes; a number of substructures representing peptides have been excluded. Short peptides were included in the data set and thus peptide nodes were incorporated into the tree. For example, node 11 represents a class of short peptides containing phenylalanine, and node 25 represents those containing proline. However, the number of peptides in the DNP is very low and therefore there is not a fair comparison between the databases. Although short peptides were included in the original data set, they will not be considered in the results, as this method of structural analysis is deemed unsuitable.

As a very general rule, child nodes contain less members and more complex substructures than their parents. The substructures given in Figure 3 all contain rings, be it aromatic, heterocyclic or a simple ring. This may be important, but alternatively may reflect the large number of molecules in the bitter database containing ring structures. A small number of nodes, specifically number 3 from Figure 3, contain 100% bitter molecules; thus none of DNP molecules contain the node substructure. Although this may suggest that the node substructure is important in bitterness, in this instance this is unlikely to be the case. Most of the bitter members of this node originate from one study (Gienapp and Schröder, 1975) in which analogues were synthesized for testing. It is therefore not surprising that there are no DNP molecules containing this substructure. In order to include as many structures as possible in the bitter dataset, synthetic molecules have not been excluded. Therefore, the source of bitter molecules must be considered when examining node members.

The substructures representing the nodes that contain a significant proportion of bitter members should be considered independently. As there are many receptors through which bitterness is mediated, there will be many different substructures that are able to activate them. The structural diversity of molecules able to produce a bitter taste indicates that no one single substructural group is able to confer bitterness.

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

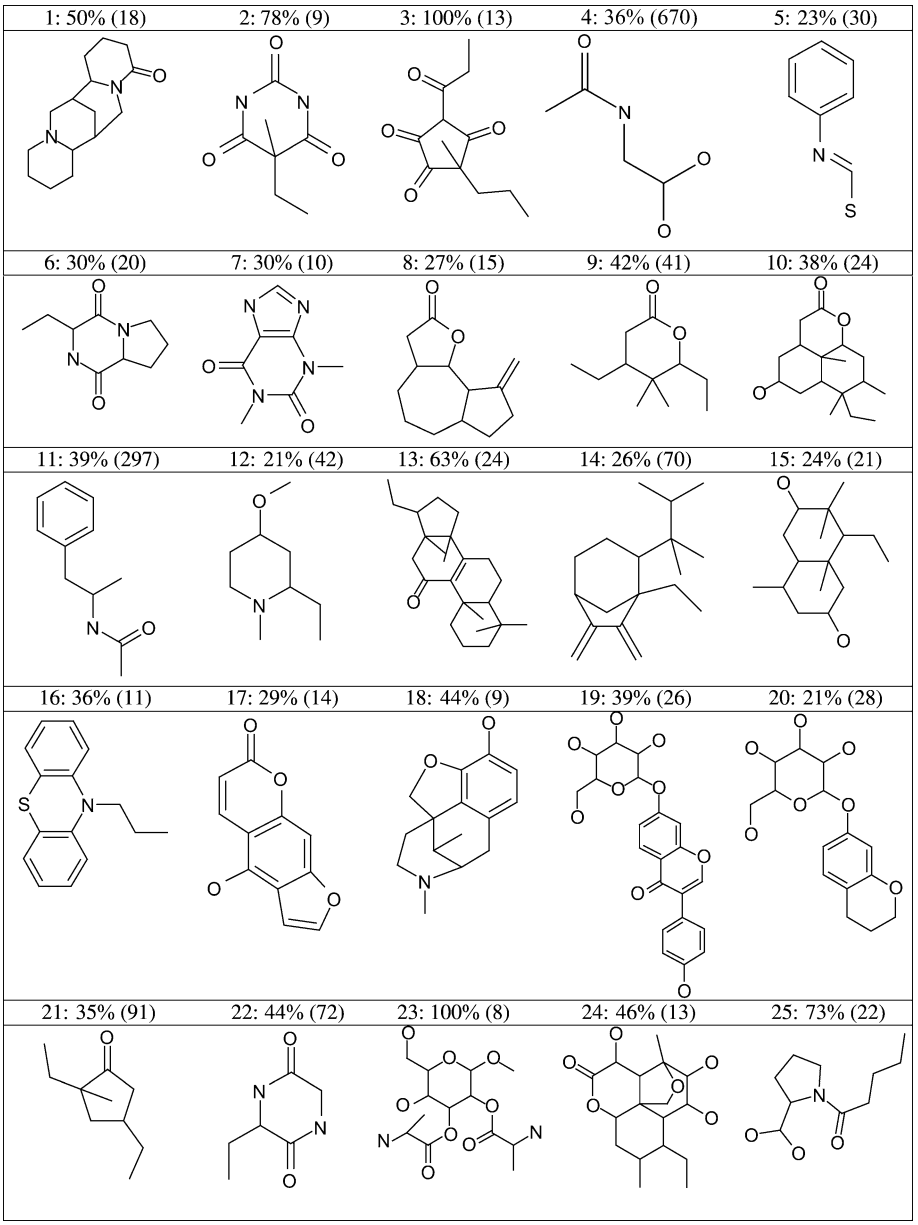


Figure 3 Significant node substructures [*node number: %bitter compounds (total membership)*].

Singletons

Seventy-five of the 833 bitter molecules contain none of the tree's substructures and therefore have no node membership in the tree. These molecules can effectively be considered singletons; they have unusual structures and share little similarity with the other bitter molecules. Nodes are only created where a few molecules share a common substructure; thus, molecules with unique or unusual structures are not included in the analysis unless they have structural features common to a number of other bitter molecules.

Four of the singletons from the tree are shown in Figure 4. Many of the singletons are fairly small molecules; larger structures have an increased likelihood of containing one of the node substructures. Valine does not have a particularly unusual structure, but contains none of the substructures present in the PGLT. In addition, the amino acids alanine, cysteine and leucine contain none of the tree's substructures. Although a number of peptide-related nodes are present in the tree, these tend to include the peptide bond, which is absent in the amino acids. In contrast, acesulfame has an unusual structure and thus shares little structural similarity with the tree's nodes. However, acesulfame is important as it is a known ligand of the human TAS2R43 and TAS2R44 bitter receptors (Kuhn *et al.*, 2004). Other singletons include a small number of glycol ethers, amines, diols, acetamides, the antibiotic streptomycin and the sweetener palatinit.

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

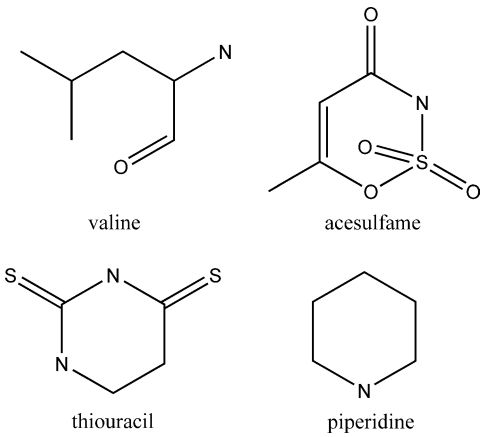


Figure 4 Singletons from the PGLT.

The singletons were clustered to identify any additional structural groups that may be present. One small group was identified and added to the top layer of the tree. The remaining singletons do not share enough two-dimensional structural similarity with other bitter molecules to join any of the existing nodes, or to form new nodes. The singletons may be important bitter tastants, but little knowledge can be gained concerning their relationship between structure and activity (bitterness) using the PGLT method.

R-Tables

R-Tables can be derived from the PGLT; an R-table details the effect on activity (in this case bitterness) of changing the substituents (or R-groups) attached to a common core structure. The tables were created from the individual nodes in the tree; the substituents of both bitter and random molecules are recorded to highlight structure–activity relationships within a given class. One such (summarized) R-table for a class of psoralens (node 17) is given in Figure 5, along with the common substructure that defines the node. The psoralens are present in grapefruit juices. They tend to be present in higher concentrations earlier in the season and thus are associated with an immature flavour (Berry and Tatum, 1986).

This node contains 29% bitter molecules, with a total of 14 molecules from the bitter and DNP data sets containing this substructure. This proportion is considered significant and suggests the substructure representing the node compounds is associated with causing the bitter taste in some molecules. All but one of the bitter node members contain a hydrophobic chain at R1, varying alkyl chains are connected through a prenyl group to the oxygen atom. None of the random molecules contain a prenyl group at R1, and there is more variation than on the bitter molecules with some consisting of just a carbon atom and others containing sugar groups. In addition, many of the random molecules contain a group attached at R2. This suggests that this group may contribute to preventing binding to the bitter receptor, as none of the bitter molecules contain a substituent here. These molecules may require a hydrophobic portion, in the form of the R1 group, in addition to the psoralen ring structure for association with groups on one or more of the bitter receptors.

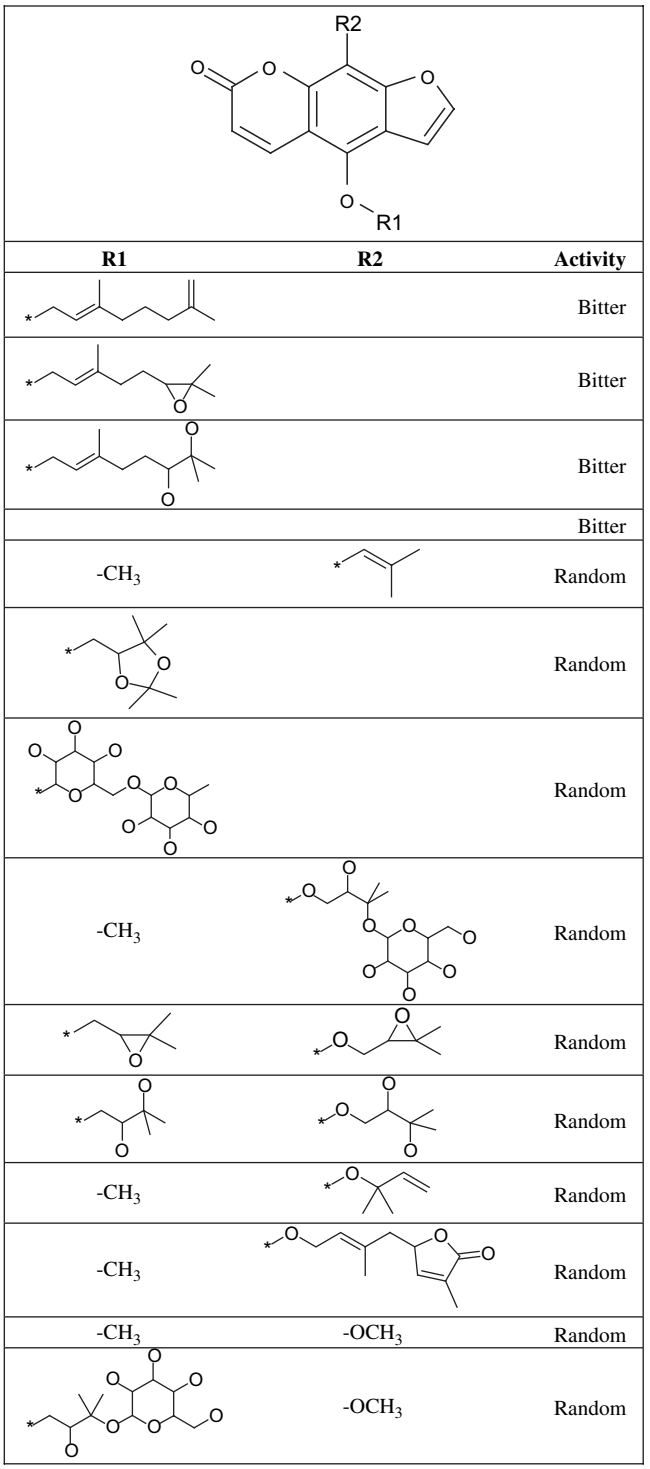


Figure 5 R-Table for psoralen class (*indicates point of attachment to core substructure where a line graph is used).

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

Multi-domain analysis

The PGLT incorporates the multi-domain nature of many molecules in that each molecule can be a member of all of the nodes with which it shares a common substructure. Thus, all the structural components on a particular molecule are being considered and information about each can be extracted. In addition, this allows for the identification of the smaller or under-represented chemical groups, as all molecules containing a specific substructure are included regardless of what other substructures they contain.

Hesperidin, a flavonoid present in many citrus fruits, provides an example of a multi-domain compound. This molecule was found to contain eight of the substructures that form the nodes of the tree, and a selection of these are given in Figure 6. Of the eight node substructures contained within hesperidin, only one contains a significant proportion of known bitter molecules.

The substructure representing node 20 is an assemblage of many of the substructures from the other nodes. A significant proportion of the molecules containing this substructure is known to be bitter. Individually, or where just two of the groups are present, the sugar moiety, benzene ring and oxane ring are not important in bitterness; however, linking the three groups together—as occurs in node 20—produces a substructure that can be considered significant. The relationship between structure and taste of the flavonoids has been well studied (e.g. Rouseff, 1980). No single structural moiety on these molecules has been identified as accounting for the bitter taste and it is likely that more than one is involved. However, the link between the two sugar groups has been identified as very important, as has the positioning of substitutions on the rings. The node substructure presented here may also represent a key component required for bitter taste.

Functional groups

A general examination of the presence of particular structural moieties within the different node substructures can highlight those likely to be of importance in bitterness. Lactone rings were found to be present in six of the significant node substructures from the PGLT; two of these substructures are five-membered rings and the others six-membered (one of which contains a double bond). The substructures are given in Figure 7, with the lactone ring highlighted in bold. All but one (node 30) of the lactone-containing nodes contain a high proportion of known bitter molecules, ranging from 27 to 46%; these proportions are given in Table 1. The total sum of members in the different nodes varies a great deal, thus care must be taken in interpreting these results. However, the total number of molecules being considered is fairly large, due to the occurrence of lactone in many different node substructures.

Due to the multi-domain nature of the PGLT, some molecules have membership in more than one lactone node. Thus to gain an overall proportion of bitter molecules containing the lactone ring only the unique molecules are considered. Where this is the case, 10% of the molecules containing a lactone ring are known to be bitter. However, this is largely due to the high number of random molecules containing five-membered lactones. If only six-membered rings are considered, then the proportion is 66%, which is clearly significant. This suggests that the six-membered lactone ring is associated more with bitterness than the five-membered counterpart.

The presence of random molecules in the lactone nodes indicates that some lactone-containing structures do not produce a bitter taste. The lactone group may be associated with bitterness, but additional components on the molecules may enhance or inhibit the biological activity, resulting in different taste characteristics. For example, the psoralens (Figure 5) all contain a lactone group, but the bitterness of these compounds depends on which R-groups are attached. This analysis suggests that *in general* the presence of a lactone group in a molecule can be associated with a bitter taste.

Discussion

The present study offers an alternative approach to the characterization of bitterness, and the methods followed may be used for the analysis of other data sets. Bitter molecules are classified according to whether they produce a bitter taste; human perception of this diverse range of molecules is the same, independent of which bitter receptor the signal is mediated through (Adler *et al.*, 2000). Although ligands for a few of the bitter receptors have been identified (Bufe *et al.*, 2002; Behrens *et al.*, 2004; Pronin *et al.*, 2004), most known bitter molecules have not been associated with a specific receptor. Collective examination of the molecules identified as producing a bitter taste response has enabled the identification of key structural groups and possible structure–activity relationships.

The database of 833 bitter molecules is used to represent existing bitter molecular structures; nevertheless, it is impossible to calculate how effectively, in total numbers and diversity, it achieves this. This limits the scope of the study; however, the database provides a fair representation of the bitter molecules that have been reported and thus those that can be considered important or interesting. Due to the large range of sources of the molecules in the database, it is likely that there will be false positives present; this is a limitation of obtaining information in this way. However, the PGLT is, to a degree, robust against the presence of false positives, as only those molecules that share structural similarity with other molecules are included in the tree.

The inclusion of a dataset of random molecules is key to identifying those structural features that are able to distinguish bitter molecules. Ideally, a data set of known non-bitter molecules would have been used to enable the identification of specific differences between bitter and

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

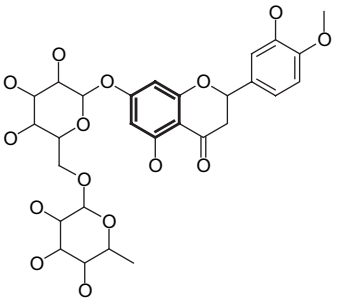
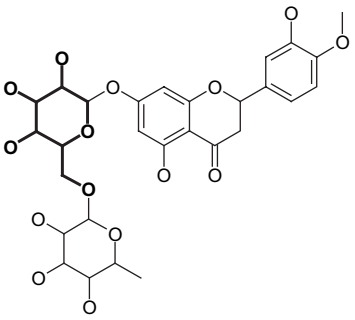
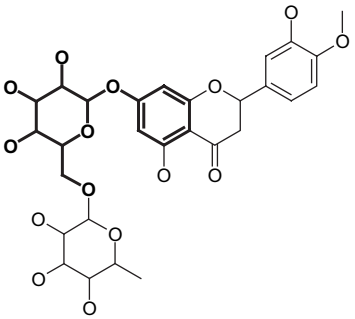
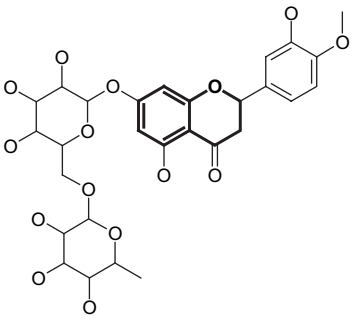
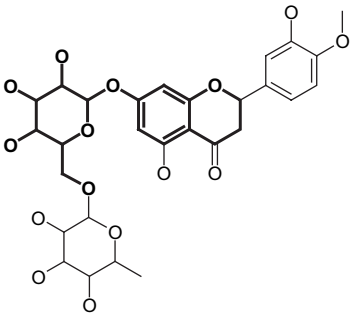
Node no.	Bitter Molecules	Bitterness (%)	Substructure
26	366	2.3	
27	108	6.3	
28	39	10.3	
29	12	3.4	
20	6	21.4	

Figure 6 Node membership of hesperidin. The node substructure is highlighted in bold.

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

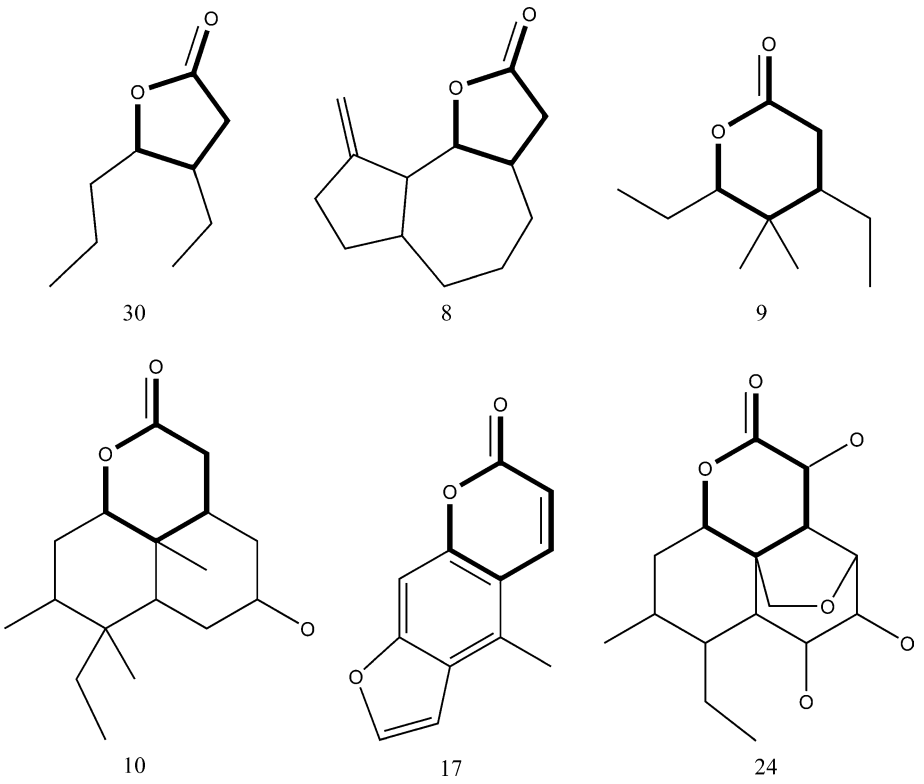


Figure 7 Lactone-containing node substructures.

Table 1 Total and proportion of bitter molecules in lactone-containing nodes

Node no.	Total no. molecules	Bitterness (%)
30	330	4.5
8	15	27
9	41	41
10	24	38
17	14	29
24	13	46

non-bitter molecules, but this is not available. It is likely that the Natural Products test set contains some bitter molecules (false negatives), but all known bitter molecules were removed.

The PGLT generation and test data filtering is two-dimensional in nature. The PGLT method considers the connectivities between the atoms in molecules, identifying groups of molecules with atoms connected in a similar way. This limits the relationships that can be identified between structure and activity as the three-dimensional spatial arrangement of the molecules defines receptor binding. However, three-dimensional methods are not well suited to analysis of such a large diverse data set; they will become important when analysing smaller, more specific groups of molecules. The two-dimensional methods employed provide an ideal starting point to identify relationships between two-dimensional structure and activity.

This method is difficult to validate as the PGLT is providing a representation of the bitter database and representations are subjective in nature. The PGLT was employed here as an exploratory tool to investigate the molecular structures that may be associated with bitterness. As so little is understood about the structural basis of bitterness, it is difficult to measure how effectively the PGLT has identified these structural features. An alternative is to check the literature for the relationships found. The specific structural features described in relation to the psoralens and hesperidin have not previously been described in the literature; thus, these are novel relationships that should be tested. Alternatively, the lactones have been described as being bitter (Belitz and Wieser, 1985); with valerolactone (six-membered lactone) exhibiting more bitterness than butyrolactone (five-membered lactone). As described above, the PGLT identified the lactone-containing nodes to be bitter; thus for this group of molecules the relationship identified from the tree is in agreement with those determined experimentally.

Numerous cluster analyses were performed on the bitter database for the initial tree generation, to find the set of clusters that most effectively represents the data. All of the

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

structural groupings produced by these sets of clusters were present in the PGLT; the PGLT also contains substructures that were not identified through the cluster analyses. This suggests the PGLT has provided a suitable representation of the molecules from the bitter database. Furthermore, it has enabled the identification of many structural features that are likely to be important in bitterness.

As mentioned above, the results obtained from the PGLT are dependent on the data sets used to create and test the tree. Alternative random data sets were filtered through the tree to test whether the same nodes would be highlighted as significant. An alternative selection from the DNP and a dataset of molecules from the World Drug Index (WDI, 2004) were filtered through the PGLT; both new test sets were similar in size to the original DNP test set. On the whole, the same nodes originally identified as significant were highlighted in the additional tests as containing a significant proportion of bitter molecules.

The PGLT has identified structural fragments that may be associated with bitterness. Many of these substructures have not previously been linked with bitter taste and thus should be tested to determine the nature of the relationship. In particular, as different bitter receptor assays become available, different substructures can be tested as ligands, and the results compared with sensory panels. All of the structures given in Figure 3 warrant further study to determine how the molecules containing these groups exert the bitter taste. In addition, the random members of these nodes should be considered to discover how they are prevented from exerting the bitter taste. Such understanding can help in identifying natural functional ingredients that are optimized for desired health effects but minimal bitterness through the selection of required structural groups.

Additional bitter molecules that are identified can be added to the nodes by filtering through the tree; thus, as the discovery of new bitter molecules continues, the bitter tree can be updated. In addition, other molecules of interest can be filtered through the tree in an attempt to identify new relationships; for example, bitter maskers can be filtered through the tree to discover which molecules they may be able to mask. The bitter molecules from specific nodes may be used for three-dimensional modelling studies such as pharmacophore identification to characterize the spatial arrangements of the groups when they bind to the bitter receptor.

Conclusion

In summary, this report demonstrates the use of a PGLT to analyse a data set of bitter molecules. By organizing the chemical structures present in the database of bitter molecules as a PGLT, interesting chemical series have been identified and relationships between structure and bitterness have been defined. Some of the structure–activity relationships correlate with those already reported in the literature; alternatively, others have not previously been reported and thus can serve as a starting point in investigating general relationships between structure and bitterness and strategies to prevent bitterness.

Acknowledgements

We would like to thank Jos Tissen and Etzard Stolte for their comments on this manuscript. This project has been funded by a Marie Curie Fellowship of the European Community Program 'Information Systems for Rational Product Design' under contract number HPMI-CT-2002-00166. This work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl) and is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

References

Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J.P. and **Zuker, C.S.** (2000) *A novel family of mammalian taste receptors.* Cell, 100, 693–702.

Ames, B.N., Profet, M. and **Gold, L.S.** (1990) *Dietary pesticide (99.99% all natural).* Proc. Natl. Acad. Sci. USA, 87, 7777–7781.

Andres-Barquin, P.J. and **Conte, C.** (2004) *Molecular basis of bitter taste—the T2R Family of G protein-coupled receptors.* Cell Biochem. Biophys., 41, 99–112.

Bacha, P.A., Gruver, H.S., Den Hartog, B.K., Tamura, S.Y. and **Nutt, R.F.** (2002) *Rule extraction from a mutagenicity data set using adaptively grown phylogenetic-like trees.* J. Chem. Inf. Comput. Sci., 42, 1104–1111.

Behrens, M., Brockhoff, A., Kuhn, C., Bufe, B., Winnig, M. and **Meyerhof, W.** (2004) *The human taste receptor hTAS2R14 responds to a variety of different bitter compounds.* Biochem. Biophys. Res. Commun., 319, 479–485.

Belitz, H.-D. and **Wieser, H.** (1985) *Bitter compounds: occurrence and structure–activity relationships.* Food Rev. Int., 1, 271–354.

Berry, R.E. and **Tatum, J.H.** (1986) *Bitterness and immature flavour in grapefruit—analyses and improvement of quality.* J. Food Sci., 51, 1368–1369.

BIOSIS (2003) BIOSIS Previews/RN®. Philadelphia: Biological Abstracts Inc. Database is available from STN at URL http://www.stn-international.de/

Bufe, B., Hofmann, T., Krautwurst, D., Raguse, J.-D. and **Meyerhof, W.** (2002) *The human TAS2R16 receptor mediates bitter taste in response to β-glucopyranosides.* Nat. Genet., 32, 397–401.

Chadwick, D.J., Marsh, J. and **Goode, J.** (1993) *The molecular basis of smell and taste transduction.* Ciba Foundation Symposium, London.

Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S. and **Rybat, J.P.** (2000) *T2Rs function as bitter taste receptors.* Cell, 100, 703–711.

ChemDraw (2001), ChemDraw, version 7.0. Software is available from CambridgeSoft at URL http://products.cambridgesoft.com/services

Delwiche, J.F., Buletic, Z. and **Breslin, P.A.S.** (2001) *Covariation in individuals' sensitivities to bitter compounds: Evidence supporting multiple receptor/transduction mechanisms.* Percept. Psychophys., 63, 761–776.

DNP (2004) *Dictionary of Natural Products*, version 8.1. Database is available from Chapman and Hall/CRC at URL http://www.crcpress.com/

Drewnowski, A. (2001) *The science and complexity of bitter taste.* Nutr. Rev., 59, 163–169.

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012

Drewnowski, A. and **Gomez-Carneros, C.** (2000) *Bitter taste, phytonutrients, and the consumer: a review.* Am. J. Clin. Nutr., 72, 1424–1435.

Froloff, N., Faurion, A. and **MacLeod, P.** (1996) *Multiple human taste receptor sites: a molecular modeling approach.* Chem. Senses, 21, 425–445.

FSTA (2003) Food Science and Technology Abstracts. Shinfield: IFIS Publishing. Database available from: Ovid Technologies Inc. at URL http://www.ovid.com/site/index.jsp

Gienapp, E. and **Schröder, K.L.** (1975) *Struktur und Bittere in der Hulupones reihe [Structure and bitter taste in the hulupone series].* Food/Nahrung, 19, 697–706.

Hartigan, J.A. and **Wong, M.A.** (1979) *A K-means clustering algorithm AS136.* Appl. Statist., 28, 126–130.

Hatzold, T., Elmadfa, I., Gross, R., Wink, M., Hartmann, T. and **Witte, L.** (1983) *Quinolizidine alkaloids in seeds of* Lupinus mutabilis. J. Agric. Food Chem., 31, 934–938.

Ishibashi, N., Kouge, K., Shinoda, I., Kanehisa, H. and **Okai, H.** (1988) *Studies on flavoured peptides. 5. A mechanism for bitter taste sensibility in peptides.* Agric. Biol. Chem. Tokyo, 52, 819–827.

Jaccard, P. (1901) *Étude comparative de la distribution florale dans une portion des Alpes et des Jura. [Comparative study of the distribution of flora in a region of the Alps and the Jura]* Bull. Soc. Vaudoise Sci. Nat., 37, 547–549.

Johnson, M.A. and **Maggiora, G.M.** (1990) Concepts and applications of molecular similarity. Wiley, New York.

Keast, R.S.J. and **Breslin, P.A.S.** (2002) *Cross-adaptation and bitterness inhibition of L-tryptophan, L-phenylalanine and urea: further support for shared peripheral physiology.* Chem. Senses, 27, 123–131.

Kier, L.B. (1972) *Molecular theory of sweet taste.* J. Pharm. Sci., 61, 1394–1397.

Kuhn, C., Bufe, B., Winnig, M., Hofmann, T., Frank, O. and **Behrens, M.L.** (2004) *Bitter taste receptors for saccharin and acesulfame K.* J. Neurosci., 24, 10260–10265.

Lee, S.J., Kim, J.-C., Kim, M.J., Kitaoka, M., Park, C.S., Lee, S.Y., Ra, M.-J., Moon, T.W., Robyt, J.F. and **Park, K.H.** (1999) *Transglycosylation of naringin by* Bacillus stearothermophilus *maltogenic amylase to give glycosylated naringin.* J. Agric. Food Chem., 47, 3669–3674.

Meiselman, H.L. and **Halpern, B.P.** (1970) *Effects of* Gymnema sylvestre *on complex tastes elicited by amino-acids and sucrose.* Physiol. Behav., 5, 1379–1384.

Nicolaou, C.A., Tamura, S.Y., Kelley, B.P., Bassett, S.I. and **Nutt, R.F.** (2002) *Analysis of large screening data sets via adaptively grown phylogenetic-like trees.* J. Chem. Inf. Comput. Sci., 42, 1069–1079.

Pipeline Pilot (2004) Pipeline Pilot, version 3.0.5.0. Software is available from SciTegic Inc. at URL http://www.scitegic.com/

Price, K.R., DuPont, M.S., Shepherd, R., Chan, H.W.S. and **Fenwick, G.R.** (1990) *Relationship between the chemical and sensory properties of exotic salad crops—coloured lettuce* (Lactuca sativa) *and chicory* (Cichorium intybus). J. Sci. Food Agric., 53, 185–192.

Pronin, A.N., Tang, H., Connor, J. and **Keung, W.** (2004) *Identification of ligands for two human bitter T2R receptors.* Chem. Senses, 29, 583–593.

Robichaud, J.L. and **Noble, A.C.** (1990) *Astringency and bitterness of selected phenolics in wine.* J. Sci. Food. Agric., 53, 343–353.

Rouseff, R.L. (1980) *Flavonoids and citrus quality.* ACS Symp. Ser., 143, 83–108.

Ruiz, M.A. and **Sotelo, A.** (2001) *Chemical composition, nutritive value, and toxicology evaluation of Mexican wild lupins.* J. Agric. Food Chem., 49, 5336–5339.

Ruiz-Avila, L., Ming, D. and **Margolskee, R.F.** (2000) *An* in vitro *assay useful to determine the potency of several bitter compounds.* Chem. Senses, 25, 361–368.

Schiffman, S.S., Booth, B.J., Losee, M.L., Pecore, S.D. and **Warwick, Z.S.** (1995) *Bitterness of sweeteners as a function of concentration.* Brain Res. Bull., 36, 505–513.

SciFinder (2004) SciFinder 2004 edition. Software available from American Chemical Society at URL http://www.cas.org/SCIFINDER/scicover2.html

Shallenberger, R.S. and **Acree, T.E.** (1967) *Molecular theory of sweet taste.* Nature, 206, 480–482.

Suzuki, H., Kajimoto, Y. and **Kumagai, H.** (2002) *Improvement of the bitter taste of amino acids through the transpeptidation reaction of bacterial γ-glutamyltranspeptidase.* J. Agric. Food Chem., 50, 313–318.

Tamura, S.Y., Bacha, P.A., Gruver, H.S. and **Nutt, R.F.** (2002) *Data analysis of high-throughput screening results: application of multidomain clustering to the NCI anti-HIV data set.* J. Med. Chem., 45, 3082–3093.

Tancredi, T., Lelj, F. and **Temussi, P.A.** (1979) *Three-dimensional mapping of the bitter taste receptor site.* Chem. Sens. Flav., 4, 259–265.

Unity (2004) Unity, version 4.41. Software is available from Tripos Inc. at URL http://www.tripos.com

WDI (2004) *World Drug Index.* Database is available from Derwent Publications Ltd. at URL http://thomsonderwent.com

WPIDS (2003) *Derwent World Patents Index Subscribers file.* Thomson Scientific, Philadelphia. Database is available from STN at URL http://www.stn-international.de/

Accepted July 12, 2005

Downloaded from http://chemse.oxfordjournals.org/ by guest on October 3, 2012